Fuzzy Rule Based Diagnostic System For Detecting The Lung Cancer Disease

Javad Hamidzadeh 1, Reza Javadzadeh*2, Ayoob Najafzadeh 2

1Faculty of Computer Engineering and Information Technology, Sadjad University of Technology, Mashhad, Iran
2Department of Computer Engineering, Azad University of Ferdows, Ferdows, Iran
*Corresponding Author

ABSTRACT

Lung cancer is responsible for the most cancer deaths in both men and women throughout the world. There are different factors for different types of cancers. Lung cancer is a type of cancer that starts in the lungs. The purpose of this paper is to design a fuzzy expert system for the diagnosis of lung cancer. The system has 9 input parameters and 1 output, in which the inputs are “age”, “smoking”, “persistent cough”, “coughing up blood”, “hoarseness of voice”, “bone pains”, “chest pain”, “shortness of breathee”, “weight loss” and the output parameter which is based on the advances and spreading of tumor, is divided into 3 stages. we have used mamadani infernece engine to deduce from the input parameters to stage the cancer. results of the experiments have been compared with state-of-the-art methods which show superiority of the proposed method in terms of detection accuracy of staging lung cancer. The superiority reason of our system is based on considering important and common symptoms of lung cancer, and removing irrelevant and possible bronchitis and pneumonia symptoms, and also we have considered the duration of time which symptoms started to pose themselves.

Keywords: Risk Analysis, Preliminary Detection, Lung Cancer, Fuzzy Logic, Artificial Intelligence

1. Introduction

Lung cancer, also known as carcinoma of the lung or pulmonary carcinoma, is a malignant lung tumor characterized by uncontrolled cell growth in tissues of the lung. In this paper, we present a novel fuzzy system to detect the stage of lung cancer considering important and common symptoms of the disease and duration time the symptoms started. Systems designed to address the issue of detecting lung cancer had a poor quality design due to considering many irrelevant symptoms, Or by concentrating on the use of better algorithms than using more trustworthy and more accurate data set. Our proposed fuzzy system is able to detect the lung cancer in patients with more precision than all previous systems which attended to address the issue. In our fuzzy system, staging the lung cancer disease, has been conducted according to the specific symptoms associated with the specific stage of cancer, for detecting the stage with more precision.

Nowadays, fuzzy system are being extensively used in different types of medical systems. In [1] a Fuzzy rule based mamadani-type inference system designed which simulates the opinion of specialists with the high accuracy of 96 percent in assessment of breast cancer risk in person. In [2] a fuzzy system using visual studio and using weighted average has been designed, but also [2] is considered too much symptoms which most of them are related very much to pneumonia and bronchitis (e.g. fever) which is going to make the performance suffer. In [3] system is using PCA for dimensionality reduction of features in the dataset, and also it is using genetic algorithm and simulated annealing to decrease the rules for decision making. in [4] designed fuzzy system can list the lung related diseases based on the probability of being exposed, ascendingly.

In [5] fuzzy system uses the same designed algorithm in [2], and output, which is based on the advances if the disease comprised 6 stages, from not being a cancer to life critical stage.

Although a variety of of lung cancer detection systems have been proposed in the literature, no single approach can be considered superior or a guarantee for satisfactory results in terms of classification accuracy or efficiency.
The purpose of this paper is to detect the stage of cancer and to do so, we used the common and important symptoms as the most influential symptoms in detection of stage of lung cancer. We also used, the duration of symptoms being started as one of the most important parameters in the detection of this disease, which was never used before in existing related papers.

The rest of the paper is organized as follows: in continuing of Section 1, we briefly introduce fuzzy expert systems and we will analyze and compare the related methods with our proposed system. In Section 3, we will explain the designed system and we will describe fuzzy functions and we will show some of used rules in the proposed fuzzy system. In Section 4, experimental results of the proposed system presented. In Section 5, we summarize our conclusion and we will propose combining our methods to get even better accuracy.

1. 1. Diagram Of Designed Fuzzy System

In Figure 1 the structure of designed fuzzy system has been shown

![Diagram of Designed Fuzzy System](image)

**Figure 1.** Structure of designed fuzzy system

1. 2. Lung Cancer Disease

Lung cancer is the abnormal growth of cells in one or both lungs, and usually occurs in cells that are in the path of air exchange. Abnormal cells won’t produce healthy normal tissues, they will divide rapidly and form tumors. Detection of lung cancer in early stages is significantly important in treatment. In America, 1 of every 14 men, and 1 of every 16 women are having lung cancer. The amount of risk for smoking people is even more. Most of the patients of the lung cancer are diagnosed at the age of 60 [6]. There are 2 different types of lung cancer, of this too, small cell lung cancer is more aggressive, it means that in the early stages it can spread faster to the other organs of the body. Small cell lung cancer is related very much to smoking, and is rarely found in non-smokers. Non-small cell lung cancer is growing slower in the body, but it is more common, and it covers almost 90 percent of all the lung cancer cases.

There are 4 classes for staging the lung cancer, but, in sum, staging is divided into 3 general stages of the disease:

- **Stage 1**: the area is small and confined to the lung and 85 to 90 percent of the time it’s treatment with surgery is enough.
- **Stage 2**: area is small, and it includes the lungs and lymph nodes and treatment is chemotherapy or chemotherapy combined with radiation therapy and surgery.
- **Stage 3**: area is larger and in addition to lungs and lymph nodes, it has spread to other organs and treatment is chemotherapy and sometimes chemotherapy combined with radiation therapy [7].

2. SURVEY OF LUNG CANCER DETECTION SYSTEMS

In [2], the system will get symptoms of the patient, each symptom according to its own impact has its own priority value which has a direct impact on it’s membership degree. A new algorithm for
inference rules are designed so that the output of the algorithm is the stage of the disease which divided into 3 parts “limited”, “extensive”, “stage 4”. At the end of the paper, the quality phrase “good” is specified for describing the performance of the system. In [3] designed system, first of all, starts to learn the given dataset, it uses PCA for dimensionality reduction and decreasing the considered symptoms from 56 to 6 symptoms and it uses genetic algorithm and simulated annealing together, to reduce the number of rules in rules database. The disadvantage of this system is that the time complexity is very high. In [4] a fuzzy system designed to detect 19 diseases related to the lungs including malignant lung cancer. The system uses Meta-Rules, which, according to it, the user enters the symptoms, and accordingly system will decide which diseases it should analyze, and that’s how removes irrelevant rules, and as a result show’s the user related probable lung diseases. In [5] designed fuzzy algorithm uses the same designed algorithm in [2], and output based on progress of cancer and so is divided into 6 sections: “no cancer”, “potential possibility”, “low stage”, “medium stage”, “high stage” and “critical stage”.

The main differences between the proposed method and each of the previous versions are

1) Considering “Age” and “Smoking” in diagnostic of the system for the first time, as they are very important parameters, to increase the accuracy of detecting the lung cancer in very early stages.
2) Removing irrelevant symptoms which would cause the confusion in detecting lung cancer, as they are very much related to many lung diseases.
3) Considering “The duration of symptoms”, which can be a key in detection of lung cancer in very early stages, and it is also a significant parameter in detection of lung cancer in other stages of the disease.

3. THE PROPOSED METHOD

In this section, we propose a novel Mamdani-type fuzzy inference system for the detection of lung cancer stages. In our proposed system, we use the most crucial symptoms of lung cancer, based on the fact that they will occur much more frequently than any other symptoms, and also, we have not considered the irrelevant and possible bronchitis and pneumonia symptoms, which was the case in most other previous works, into our system. There is a relation between the occurrence of some of the symptoms and the growth in stage of the lung cancer into stage 2 and stage 3. And given the critical importance of early detection of lung cancer, we also used the number of days that the symptoms have emerged as an input parameter. If considered symptoms last more than 2 to 3 weeks it can be a critical sign of lung cancer.

3. 1. Determining Membership Functions

We used 9 input parameters, and also “age” and “smokingAndExposure” to assess the possibility of people having lung cancer in case of low duration time of symptoms. In the following, we will discuss the symptoms and their corresponding membership functions.

3. 1. 1. Age

Most lung cancer cases occur in people over 60 years old. Because lung cancer takes many years to reach the stage where the patient will feel the symptoms and visits a doctor.
3. 1. 2. Smoking and exposure to carcinogenic gases

As noted, smoking is directly related to lung cancer. 2/3 of patients with lung cancer are smoker. Also, people who are regularly exposed to tobacco smoke, and asbestos factory workers, foundry workers, are much more than the others at risk of lung cancer due to exposure to carcinogens gases or materials [8]. the input consists of two parts, yes or no.

3. 1. 2. Persistent Cough

Measurement unit in this input, is the number of days that person suffers from persistent coughs which has not gotten better.

3. 1. 4. Coughing Up Blood

Measurement unit in this input, is the number of days that person suffers from persistent bloody coughs which has not gotten better. This symptom, for the sake of being more important that other symptoms, if it takes more than 10 days, it enters into medium membership function.
3. 1. 5. Chest Pain

Measurement unit in this input, is the number of days that a person suffers from chest pain. Because most patients with lung cancer visiting a doctor will take more than a week and because this symptom in patients occurs almost late, it enters into medium membership function sooner than some other symptoms.

3. 1. 6. Weight Loss

If there is more than 10 percent weight loss less than a month, with the presence of other symptoms it is a sign of lung cancer.

3. 1. 7. Shortness Of Breath

If it takes more than 2 to 3 weeks it can be a sign of lung cancer [9]. 20 to 90 percent of cases occurs in advanced stages of cancer [10]. Measurement unit in this input, is the number of days that a person suffers from shortness of breathing.
3. 1. 8. Bone Pains

The aim of bone pain, is more of a back pain, and pain in the waist and knees. This factor occurs when the tumor has spread to the bones, hence the cancer is in advanced stages. It should be noted that for patients with bone disease this parameter should take the value 1, hence not considering this parameter for patients of bone diseases.

3. 1. 9. Hoarseness Of Voice

Measurement unit in this input, is the number of days that a person suffers from hoarseness of voice. This symptom is obvious during speaking of a patient.

3. 1. 10. Output

Divided by 4, which is "no cancer", "Lung Cancer - Step 1" Lung Cancer - Step 2" and "Lung Cancer - Step 3", Our proposed system is able to determine the stage of lung cancer depending on the severity of symptoms.
3. 2. Discussion

In Lung cancer, some symptoms like “hoarseness of voice” mostly occurs in stage 2 so it also is one of the symptoms in stage 3. some symptoms like “unintentional weight loss” is much more likely to happen in stage 3 than in stage 2 of the cancer, and also “shortness of breathe” is one of the most important symptoms which often occurs in stage 2 or 3, However, “bone pains” most oftenly occurs in stage 3 which is the most advanced stage and, the tumor has spread to the bones, The presence of small levels of these symptoms along with the other symptoms like “shortness of breath”, “coughing up blood”, “persistent cough” and considering that the person is a smoker and aged, there is a strong possibility of existence of stage 1 lung cancer, in that person. We have considered all of these situations in designing our system.

Lung cancer symptoms often detected in stage of 2 or 3, which is when the tumor has spread to other parts of the body, such as the bones and lymph nodes, which, of course, at this stages, the person will have the symptoms of lung cancer. However, persistence of symptoms more than 2 to 3 weeks may be caused by lung cancer, and we've considered time limit of 2 to 3 weeks as less risky. Lung cancer is growing fast from stage 1 to advance, and unfortunately Smoking people have died during the 4 months after the onset of symptoms. In the proposed paper there has been an interview with Dr Morad Khani and the papers [11] and other papers cited in the references have been used. And also we used the video interview of several specialists and researchers. And we also explored in the symptoms of lung cancer patients, according to the video interviews of the patients.

3. 3. Fuzzy Rule Database

Fuzzy rules and fuzzy inference systems are extremely important and the quality of the system very much depends on them. These rules are defined as “if-then” rules. In this system, we used 81 “if-then” fuzzy rules, as following examples:

- **Rule 1**: If (persistent cough is very high) and (coughing up blood is very high) and (chest pain is very high) and (weight loss is very high) and (shortness of breath is very high) then (lung cancer-stage 3)

- **Rule 2**: If (persistent cough is low) and (coughing up blood is low) and (chest pain is low) and (weight loss is low) and (shortness of breath is low) and (horseness of voice is low) then (no-cancer).

- **Rule 3**: If (persistent cough is high) and (coughing up blood is high) and (chest pain is high) and (weight loss is high) and (shortness of breath is high) and (bone pain is not low) (horseness of voice is not low) then (cancer-stage 3).

3. 4. Fuzzifier and Defuzzifier
Designed system uses the mamdani inference, the defuzzifier uses the famous COG (center of gravity) method which it’s formula is as follows:

\[ y_0 = \frac{\int y \mu_B(y) \, dy}{\int \mu_B(y) \, dy} \]  

(1)

4. EXPERIMENTAL RESULTS

After determining the input parameters, output parameters are specified, which Figure 10 shows the different stages in the development process of the disease which is 4 fuzzy set output parameter (no cancer, stage 1, stage 2, stage 3). It should be noted that in general, the complete accuracy of detection based on just the physical symptoms and not using an MRI or CT-Scan information, is quite not possible. But with considered symptoms, this disease can be detected with high accuracy, hence lung cancer might be detected in early stages or at least before progressing the disease to advanced stages, thus, Getting into the treatment program as sooner as possible.

System designed in [2] does not have enough flexibility and the duration of symptoms did not counted. Symptoms in designed forms selected using the checkboxes and priority of every symptoms multiplies the value of symptom influence and is used in designed algorithm, hence, in the next form each user will be asked to choose the severity of symptoms from “very low”, “low”, medium”, “high”, “very high”. The result is, for example, as shown in Table 1.

**TABLE 1. Results Of Lung Cancer Diagnosis**

<table>
<thead>
<tr>
<th>Patient's Name</th>
<th>Cancer Type</th>
<th>Cancer Stage</th>
<th>Treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 11</td>
<td>Small Cell Lung</td>
<td>Extensive Stage</td>
<td>Chemotherapy alone or in combination with radiation therapy</td>
</tr>
</tbody>
</table>

Unfortunately, this paper is not addressing the duration of symptoms, and considers a lot of symptoms in which might be a symptom of lung cancer, but, many other factors are also causing them to appear, such as dizziness in patients with pneumonia disease. But our system by considering the most important and most common symptoms, has been able to achieve reasonable accuracy. In [3] there is 56 features but there is only 32 samples, which is very few for finding a good classifier. It should be noted that the complete accuracy is not possible without considering CT-Scan Results. The problem is that, normally, lung cancer patients won’t find out about the disease as a result of symptoms weak signs, or symptoms are hidden, or for the sake of laboratory costs; and these are going to cause the progression of the disease in advanced stages. Hence designing a system without using the collected lab data, so patients can use it anytime without or before visiting a doctor seems very imperative which our fuzzy system covers that. In [4] Lung cancer is detected just in the case of malignant lung cancer. But the diagnosis of lung cancer, in malignant stage, is not much promising for the treatment. The quality phrase “good” used to express the quality of designing the system. The output of the system, as a list of possible lung related diseases will be shown and the possibility of every one of them has been calculated for a patient. Table 2 is an example of the results for 2 patients.

**TABLE 2. The Results Of Selected Inputs**

154
This system covered the diagnosis of lung cancer in malignant state, hence, it’s not addressing the various stages of lung cancer.

In [5], The system takes the symptoms of a user and output of the system is the stage of cancer, which divided into “no cancer”, “potential possibility”, “low stage”, “medium stage”, “high stage” and “critical stage”. The system is able to detect a number of other diseases associated with lung. The result of patients’ inputs, if patient diagnosed with lung cancer, is very similar to Table 1. Form drawbacks of this system we can mention that, this system does not consider the duration of symptoms for the diagnosis of lung cancer.

Unfortunately the previous methods do not consider the onset of symptoms, or the symptoms are very much in mind, most of which may be seen in patients with other lung diseases. But our proposed method, by taking into account most common symptoms, and duration of symptoms, has gained the accuracy more than 80 percent.

In Table 3, previous methods, has been named for the sake of comparison. Table 4 shows the results of comparing the accuracy of the proposed method, with the previous methods. For the proposed method the Codename “BSH” has been used, and for the other methods:

- For [2] and [5] same dataset has been used for the detection of lung cancer, and the only difference in [5] is, it uses extended rules to diagnose other lung diseases. So they are sharing the Codename “EFR2”.
- For [3], UCI lung cancer dataset has been used, we will Codename it as “FSL1”.
- For [5], Because this method only considers the malignant stage of lung cancer, it hasn’t been used in comparisons.

It should be noted that, due to the much higher rate of patients with advanced stages of lung cancer, And normally, late appearance of symptoms, For The Method “EFR2”, we will consider:

- “Potential Possibility” and “Low Stage” as Lung Cancer - Stage 1.
- “Medium Stage” as Lung Cancer - Stage 2.
- “High Stage” and “Critical Stage” as Lung Cancer - Stage 3.

In Table 3, #samples, #features, and #classes denote the number of data samples, the number of attributes and the number of classes, respectively. Also we have to mention that, For BSH (our proposed system) and EFR2, #samples, means number of fuzzy rules.

### TABLE 3. Methods Used For Comparisons

<table>
<thead>
<tr>
<th>ID</th>
<th>Method Codename</th>
<th>#Samples</th>
<th>#Features</th>
<th># Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BSH</td>
<td>81</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>EFR2</td>
<td>72</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>FSL1</td>
<td>32</td>
<td>56</td>
<td>3</td>
</tr>
</tbody>
</table>
To test the system, 62 samples of patient profiles collected from cancer institute at ImamReza hospital in bojnord and also from patient stories in related lung cancer community websites [12, 13]. We applied these profiles to the investigated methods and our proposed method. Table 4 shows the results of the comparison. It should be noted, that the accuracy of “FSL1” method obtained, by applying this method to the patient profiles database, using the 10-Fold Cross Validation.

<table>
<thead>
<tr>
<th>Method Codename</th>
<th>Diagnosing the Stage of Cancer (%)</th>
<th>Cancer Detection - No Staging (%)</th>
<th>Average Results (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSH</td>
<td>82.25</td>
<td>87.09</td>
<td>84.67</td>
</tr>
<tr>
<td>EFR2</td>
<td>65.38</td>
<td>73.07</td>
<td>69.22</td>
</tr>
<tr>
<td>FSL1</td>
<td>79.03</td>
<td>88.70</td>
<td>83.86</td>
</tr>
</tbody>
</table>

In diagnosing the stage of cancer, the proposed method (BSH) acted better than other methods. The weakness of EFR2 was due to the weakness of the used dataset, Which reduced the accuracy of the diagnosis, especially in stage 1 and stage 2. The accuracy results of the “FSL1” is very close to our method, and the reasons why it can be, we can mention to using PCA for dimension reuction, But it should be noted that although this method is more accurate than the way we diagnose cancer without stages, But its execution speed and its computational complexity is much higher, And our system is superior in terms of computational time and overall results. The results of cancer detection without staging, represents the increase in accuracies, Because of no need to separate the stages, All 3 methods were relatively successful in separation of stage 2 from stage 3, but they were problematic in isolating cancer-free patients and stage 1 of the lung cancer. Average results demonstrate the superiority of our method in terms of classification accuracy with less running time and computational complexity than other existing methods.

5. CONCLUSIONS

In this paper, we explained the design and construction of fuzzy expert system for the diagnosis of lung cancer. For determining inputs and the membership functions and their importance we consulted with a medical specialists of lung cancer and also patient’s profiles has been used. Finally the performance of the system on the data obtained from 62 patients evaluated. Experimental results and comparison with other existing methods indicates the superiority of our proposed method, in terms of classification accuracy and computational speed. But the challenge in designing a system is in providing accurate results for detection of lung cancer in early stages to reduce the complexity of treating cancer in advanced stages. Due to high accuracy of system, Designed system can assist physicians in medical centers in assessing the risk of lung cancer in patients with possible symptoms. Designed system can also assist lung specialists in the diagnosis of lung cancer in patients. In addition, the patients can also use the system on their own. To increase the accuracy of the system, neural networks in combination with our fuzzy system can be used, also, detection of the stage of the lung cancer, using the data obtained from the results of the MRI and CT scan of the patient, can surely improve the accuracy of our designed fuzzy system.
7. REFERENCE


