

Radiography and risk factors of lung cancer: modeling using an intelligent system

Abstract

Many factors involved in the detection of lung cancer. The confirmation depends on Age; Tobacco use; Historical malignancy; Growth rate; Edge characteristics; Size). The factors analyzed in this study are not precise beings and do not induce the same effects in the groups of the population. These variables are uncertain and therefore fuzzy variables. The principles of fuzzy inference prove adequate in this case. A fuzzy system is constructed with four input variables (Age, Tobacco use, Historical malignancy, density and morphological tumor) and one output represents the rate of exactitude in diagnosis. A base rules was established. It will be possible to read the expected degree of confirmation of malign tumor or benign tumor of lung cancer by introducing randomly the values in inputs variables. The analysis of factors at diagnosis, differ for the different types of lung cancer. Several detection radiography techniques to improve sensitivity for the representation of nodules modeling in association with other factors remain imprecise. The results of statistical analysis using Bayesian approach or logistic regression are limited to significant or insignificant with degree of incertitude. After the analysis system is done, the analysis system proposed takes into consideration uncertainties in these factors.

Keywords: radiography, lung tumor, nodules, fuzzy logic

Introduction

The fidelity of the image obtained in radiography remains always without much success. While different techniques are used for the purpose of its clarification. Parasitic effects are to be eliminated such as the effects of bones. Generally, imaging analysis is necessary when it comes to nodule detection using X-ray. In many cases, nodules detected on radiography require other assessment techniques in addition to imaging.¹ Differential diagnosis is often necessary when a solid pulmonary nodule is detected. To make the difference it is essential to evaluate the evolution of the SPN.

To have an idea on the evolution of the nodule in volume terms is necessary to fix itself on the image obtained in imagery. Recent techniques of tomography scans allow more frequent detection of pulmonary nodules. The difficulty in the case of an undecided diagnosis is the determination of the nature of the neoplastic lesion. In order to have a clear and precise idea of the nature of the tumor, consideration must be given to the effects of external factors in relation to radiographic findings. When one is dealing with a suspicion, one must follow the evolution of the size to better guide the diagnosis.² Based on several clinical trials and different results, modeling techniques evolve of the repeat cases were recorded and from their prevalence follow-up.³ For example, technical personnel qualified in radiological reading are lacking is a handicap to tuberculosis screening in countries where tuberculosis is endemic. Nowadays digital techniques have become indispensable for radiographic analysis. Current techniques that support clinical parameters and treat them numerically do not exist.⁴ Our study demonstrates that screening assisted by fuzzy inference software can efficiently and accurately examine and sort out radiologist examinations. As there is inaccuracy in the scan, this can lead to errors of judgment. We consider that this technique is adequate when dealing with an indeterminate situation. More than the parameter number is high, more than the result is

refined. The uncertainty and imprecision associated with the very nature of the biological parameters is palliated by fuzzy fication.

Risk factors

Clinical features are independent predictors of malignancy include age, smoking and historical pulmonary diseases. Age is a factor that increases the probability of the malignant nature of the pulmonary nodule. Another factor to consider is smoking. In addition, the history of nodule recordings can predict a malignant nodule.

Age

No doubt that lung cancer is directly related to age. When it comes to comparing different groups, age is the basis of this especially with regard to lung cancer. Risk increases with age.

Sex

It has been proven that the type of lung cancer differs from one sex to another in terms of frequency. Men for example have the greatest risk of squamous cell carcinoma than men and with lower adenocarcinomas. Histologically, this may explain these differences in risk rates.⁵ In general, continuous pulmonary carcinomas dominate in men than in women. Some carcinomas are seen observed in young women. Regardless of race or ethnicity, apart from adeoconciomas, there is a decline in men regarding squamous cell carcinomas that are directly related to tobacco. Irrespective of race, the decrease in squamous carcinomas did not decline in women than in men.⁶

Tobacco use

Irrespective of race, sex or geographic location, all epidemiological studies have demonstrated categorically that tobacco remains the most dominant factor in lung cancer. Its incidence rate is a function of the duration of the smoking and the quantity that is to say the number

of cigarettes consumed. This does not exclude that other chemical pollutants also participate in this type of cancer, such as chemical agents emanating from the professional environment or from air pollution.⁷ The difference between the tobacco-related cancer rates is 90% compared with the risk of other indirect pollutant factors.⁸ Depending on the type of lung cancer, the evolution is significant after cessation depending on the duration of smoking and the number of cigarettes smoked.⁹

History of other malignancy

The history of pulmonary nodules plays a primordial role in determining the nature of the tumor. A subject with a history, the risk of metastasis is increased. The risk of malignancy can be up to 15 times compared to a person with no history.¹⁰

Screening

Screening studies have no universally recognized techniques. However, computed tomography techniques are considered more reliable because they are more sensitive. The disadvantage is that this sensitivity can mislead and lead to false positives. This may result in additional unnecessary diagnoses for confirmation.¹¹ In the diagnosis of pulmonary nodules, its dimensions are often used as a determining parameter. The reliability of the measured dimensions takes into account certain factors, in particular those relating to the digitization and segmentation of the image obtained and the digital processing used. Depending on the technique used, accuracy depends. However, some techniques are adapted to clinical data and support different forms. Others support, in addition to volumes, the attenuation coefficients and calibrated on virtual samples. Two techniques are applicable. Those that use specific image analysis algorithms and those that are in semi-automatic mode. In terms of image processing, parasitic effects are eliminated such as surrounding matter as the vascular system.¹² In terms of height, patients without a history and having nodules (<5mm), only 1% are considered malignant. While the large nodules with dimensions (20mm) the probability of being of malignant nature is of the order of 50%. The estimated risk is defined according to the size of the nodules referring to the recommendations of the Fleischner Society.¹³ Based on these standards we establish the inference rules in terms of size and risk exposure.

Morphological tumor

Based on the morphology of the tumor, classifications are established. This is considered to be basic in classifying tumors by pathologists. Nowadays, recent techniques are used in tumor differentiation. Their importance appears especially when it comes to providing this as a tool for clinicians. Size, shape, growth rate and density are considered to be the main parameters of tumors.¹⁴ Differentiation between true and false nodules remains problematic despite the use of digital imaging and computational analysis techniques.¹⁵

Growth rate What develops between hilar tumors and squamous cell carcinomas is the rate of growth of hilar tumors.¹⁶ Depending on the surface condition of the nodules, these are classified as smooth or spiculated. What differentiates small cell carcinomas from non-small cells is also the rate of growth with a high rate and rapid development of metastases as well as their responses to chemotherapy and radiotherapy.¹⁷ In our study, this factor is not taken into consideration.

Density Among the parameters taken into consideration in the

tests, the density of the cells, in particular dendritic cells and which are mature, proves to be a reliable indicator of prediction. Depending on the number of tumors entering the tumors, affected patients may identify themselves early as relapsing.¹⁸ In this study, we improved the lung nodule detection system by using fuzzy inference system gained from radiologists such as anatomical characteristics and morphological features of lung nodules (size, and shape) and other risk factors (Age, sex, tobacco use and History of other malignancy). A fuzzy algorithm system may improve the sensitivity of cancer detection, reduce oversight errors, and decrease inter- and intra-reader variations.

Materials and methods

Statistical analysis

Bayesian approaches and logistic regression: Different statistical techniques for the analysis of data related to pulmonary tumor risk factors are used. The Bayes theorem is used in statistical inference to update the estimates of any probability from the observations and probability laws of these observations. When sufficient probabilities are available, Bayesian approaches are often used. This presupposes having already a strong knowledge of a situation. Other analysis use the logistic regression technique for identifying the factors that characterize a lung tumor of people who are sick compared to healthy people. This then makes it possible to identify the factors associated with the nature of the tumor. More precisely, logistic regression aims to isolate the effects of each variable that is to identify the residual effects of an explanatory variable on a variable of interest. These two methods remain in the probable; the result obtained is always characterized by uncertainty and expressed with a degree of inaccuracy.

Fuzzy logic inference: The factors governing the field of medicine are characterized by their uncertainty. By way of example, the age factor cannot be subdivided into sections with sharp edges. Characters related to the age of one interval, for example, can spread on the other. To overcome this imprecision, fuzzy logic proposes to analyze them in linguistic terms (young, adult and old) in terms of numerical terms. This comes close to human reasoning. A fuzzy proposition is characterized by its degree of belonging to the membership function with degrees of truth. A theory of fuzzy logic proposed by Zadeh L¹⁹ explains how to formalize "Fuzzy".¹⁹ As such, fuzzy logic can be considered as an extension of classical logic. The inference rules are of type: If X_1 is A_1 and X_2 is A_2 ... and X_n is A_n then Y is B .

These rules are based on the actual values recorded. Fuzzy logic inference has been successfully applied in different domains. Where uncertainty, complexity and imprecision are the characteristics of the system. It happens that precisely what characterize the medical field are these imperfections. Fuzzy modeling, fuzzy control and fuzzy classification are the most common applications.^{20,21} Rather than dealing with random variables, fuzzy logic treats situations where the source of data is imprecise and where different factors intervene and the exact effect of each is ignored and sometimes even when certain factors are totally ignored.²²

Fuzzy logic modeling: In our case, the factors involved in the detection of the nature of the pulmonary tumor (Age, sex, tobacco use, History of other malignancy, size, shape, growth rate and density) are characterized by their imprecision. We can introduce the relationship between an eight-inputs and an output system (Figure 1).

Fuzzy fication of variables: This first step consists to convert numerical values to linguistic variables. The inputs and output are classified into three linguistic functions.

Fuzzy fication of the input variable “Age”: The input that represents the age is expressed by three fuzzy intervals and membership functions defining the [young “0-30years”, Adult “25-55years”, Old “>50years” (Figure 2). In the same way, the other variables are fuzzy fied. The variable ‘sex’ is not fuzzy fied. We assigned a value of 1 for the male and a value of 2 for the female sex. By way of example and referring to the norms established by Fleischner Society recommendations, We attribute the size value of 0-5cm ‘low risk’, 4-8cm ‘medium risk’ and 6cm to 10cm ‘high risk’ and the nature of the tumor is confirmed malignant (Figure 3).

Fuzzy fication of the output variable “nature of tumor”: The output variable the nature of tumor is represented. It is expressed by three fuzzy intervals and membership functions defining the: benign tumor, uncertain tumor, malign tumor (Figure 4). In according with the input variables, the fuzzy rules are established. These rules are in the form: IF ‘Age’ is X₁, AND ‘Sex’ is X₂, AND ‘Tobacco use’ is X₃ AND ‘History’ is X₄, AND ‘Size’ is X₅ AND ‘Shape’ is X₆ AND ‘Growth rate’ is X₇ AND ‘Density’ is X₈ THAN ‘Nature of tumor’ is Y.

The basis of the established rules must take into account all possibilities and all possible combinations. The collaboration of the set of rules determines the value of the output. The establishment of these rules refers to the effect of each factor on the nature of the tumor.

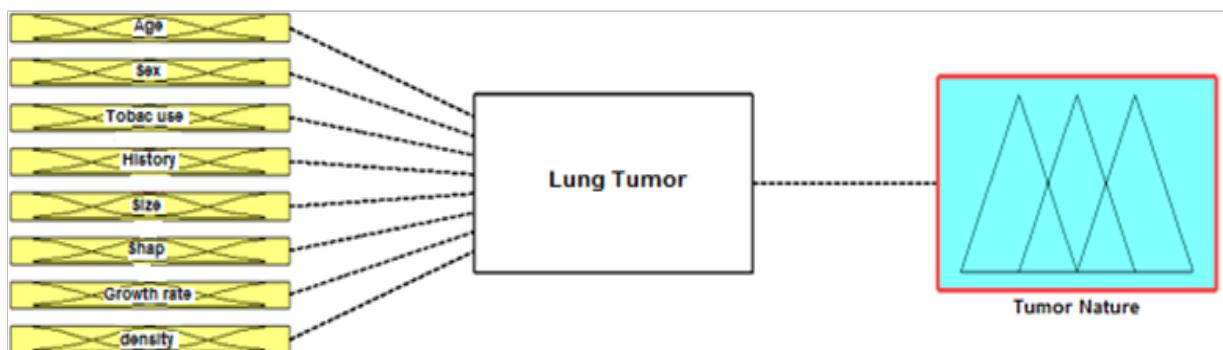


Figure 1 Block diagram of the system with eight inputs and one output.

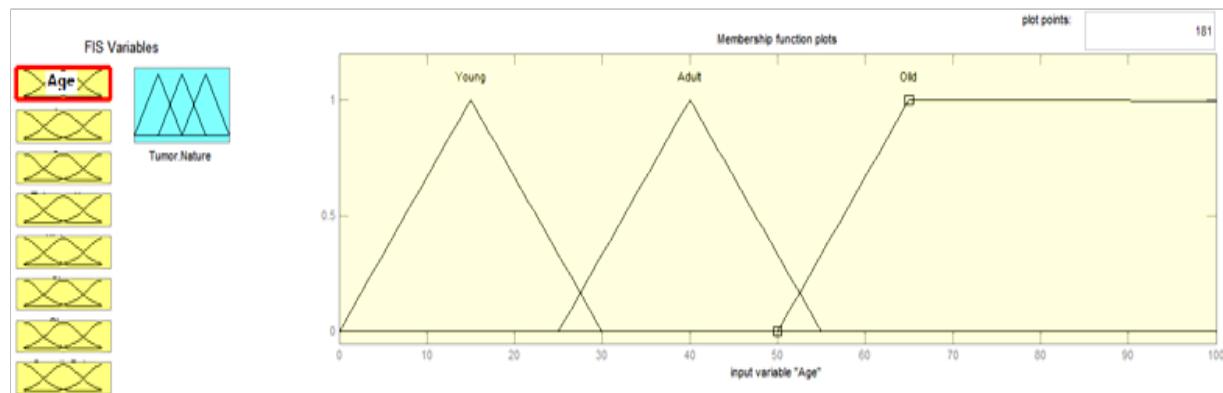


Figure 2 Values attributed to the ‘Age’ variable.

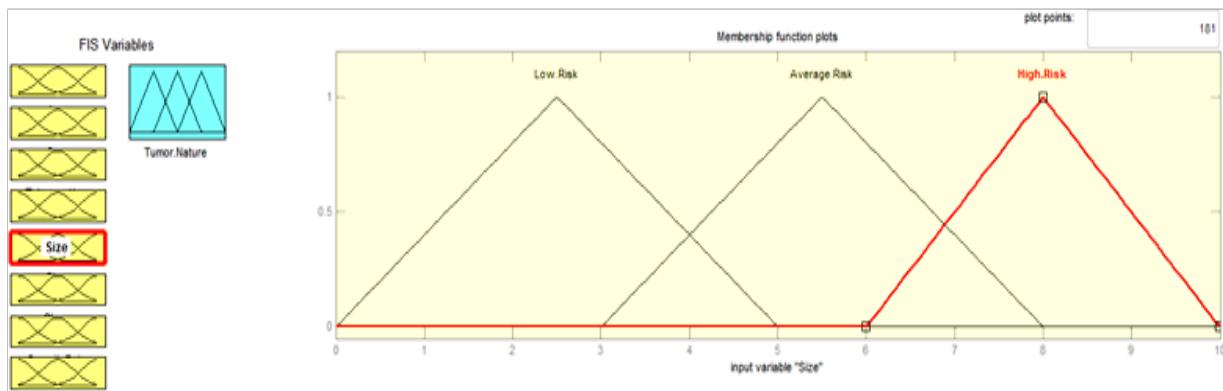


Figure 3 Values attributed to the ‘Size’ variable.

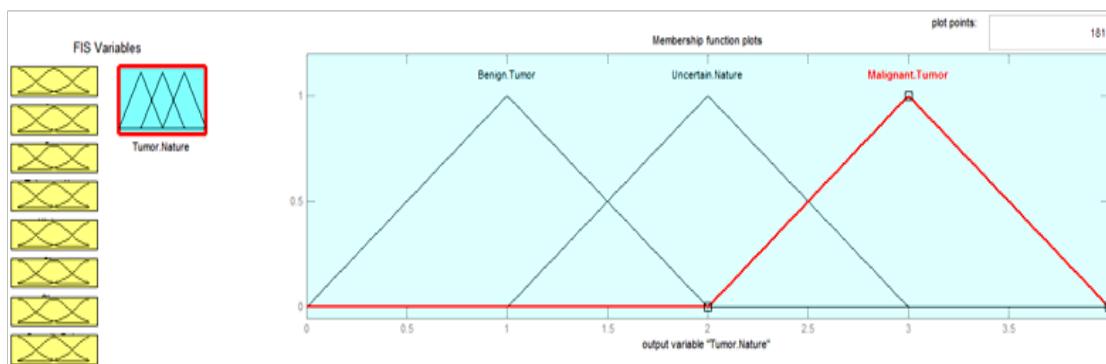


Figure 4 Values attributed to the 'Tumor nature' variable.

Conclusion

Since the factors involved in the process are characterized by complexity, uncertainty and imprecision, statistical models of analysis remain in the realm of probability. In this study, an intelligent approach is proposed. The fuzzy inference system used treats these factors as fuzzy variables. Numeric variables are replaced by linguistic variables. This comes close to human reasoning. Each parameter is fuzzy fied. This fuzzyfication takes into account these imperfections. The basis of the rules is established by a human expert. The result takes into account the collaboration of all the rules. The aggregation of rules makes it possible to give a result as precise as possible. When the rules are established correctly and encompass all possibilities, the outcome determining the nature of the tumor leaves no doubt. The

input variables considered in this study gathers the essential factors that determine the nature of the tumor. However, the proposed system remains extensible to other factors that are not supported here. The established system then makes it possible to instantly read the nature of the tumor at the output in numerical terms, attributing it to the pre-established intervals in the variable of this variable. The reading of the result at the output is possible just by entering random values at the inputs (Figure 5). This tool can in no way replace human expertise, but it can be a support tool for radiologists.

Such a fuzzy program can be incorporated into the screening radiology system where it will be possible to introduce external factors such as age, sex, antecedents, etc. by display. The parameters related to the result of the scan can be automatically taken and introduced to the systems. Then result display will be possible directly on the unit.

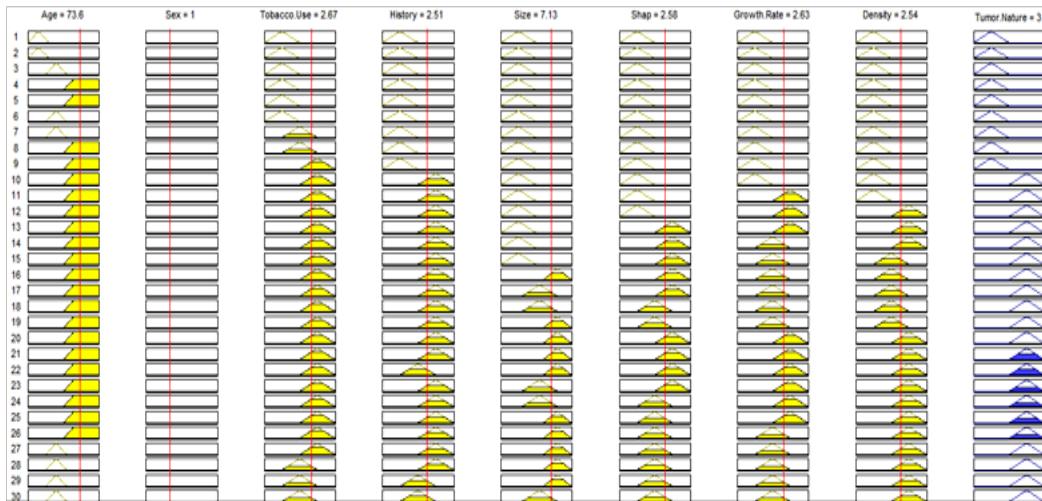


Figure 5 Reading the degree of tumor nature at the output by fixing random values at the input.

Acknowledgements

None.

Conflict of interest

Author declares that there is no conflict of interest.

References

1. Mylene TT, Jane P Ko, Santiago E Rossi, et al. Update in the Evaluation of the Solitary Pulmonary Nodule. *Radio Graphics*. 2014;34(6):1658–1679.
2. Bonetti A, Aubert JD. Le nodule pulmonaire solitaire. *Rev Med Suisse*. 2008;4:2506–2510.
3. Callister MEJ, Baldwin DR, Akram AR, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax*. 2015;2:1–54.
4. Jaime M, Clara I Sánchez, Rick HHM Philipsen, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Scientific Reports*. 2016.
5. Malik PS, Raina V. Lung cancer: Prevalent trends & emerging concepts. *Indian J Med Res*. 2015;141(1):5–7.

6. Lewis DR, Check DP, Caporaso NE, et al. US Lung Cancer Trends by Histologic Type. *Cancer*. 2014;120(18):2883–2892.
7. Jonathan MS. The Epidemiology of Lung Cancer. *Chest*. 1993;103(1):20S–29S.
8. American Lung Association Epidemiology and Statistics Unit. *Research and Program Services Division*. 2014.
9. Kenfield SA, Wei EK, Stampfer MJ, et al. Comparison of aspects of smoking among the four histological types of lung cancer. *Tob Control*. 2008;17:198–204.
10. Carlos M, Mery MPH, Anastasia N, Pappas MSW, et al. Relationship Between a History of Antecedent Cancer and the Probability of Malignancy for a Solitary Pulmonary Nodule. *Chest*. 2004;125(6):2175–2181.
11. Kris MG, Johnson BE, Berry LD, et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *JAMA*. 2014;311(19):1998–2006.
12. Marios A, Lisa M Kinnard, Kyle J Myers, et al. Noncalcified Lung Nodules: Volumetric Assessment with Thoracic CT. *Radiology*. 2009;251(1).
13. Heber MacM, John HM Austin, Gordon Gamsu, et al. Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society. *Radiology*. 2005;237(2).
14. Pelosi G, Sonzogni A, Viale G. The classification of lung carcinoma: time to change the morphology-based approach? *Int J Surg Pathol*. 2010;18(3):161–172.
15. Chiou YSP, Lure YM. Hybrid Lung Nodule Detection (HLND) system. *Cancer Lett*. 1994;77(3–25):119–126.
16. Hasegawa M, Sone S, Takashima S, et al. Growth rate of small lung cancers detected on mass CT screening. *The Br J Radiol*. 2000;73(876):1252–1259.
17. Govindan R, Page N, Morgensztern D, et al. Changing Epidemiology of Small-Cell Lung Cancer in the United States Over the Last 30 Years: Analysis of the Surveillance, Epidemiologic, and End Results Database. *J Clin Oncol*. 2006;24(28):4539–4544.
18. Dieu-Nosjean MC, Antoine M, Danel C, et al. Long-Term Survival for Patients With Non-Small-Cell Lung Cancer With Intratumoral Lymphoid Structures. *J Clin Oncol*. 2008;26(27):4410–4417.
19. Zadeh LA. Fuzzy sets. *Information and Control*. 1965;8:338–353.
20. Driankov D, Hellendoorn H, Reinfrank M. *An introduction to fuzzy control*. New York: Springer-Verlag; 1993.
21. Chi Z, Yan H, Pham T. Fuzzy algorithms: with applications to image processing and pattern recognition. *World Scientific*. 1996.
22. Bouharati S, Benmohammed K, Harzallah D, et al. Application of artificial neuro-fuzzy logic inference system for predicting the microbiological pollution in fresh water. *Journal of Engineering and Applied Sciences*. 2008;8(2):309–331.