TOWARD DETERMINATION OF VENOUS THROMBOSIS AGES BY USING FUZZY LOGIC AND SUPERVISED BAYES CLASSIFICATION

P.H. Lim[1,3], R. Debon[1], B. Solaiman[1], L. Bressollette[2], B. Guias[2], C. Roux[1], H.T. Chuah[3]

[1]Département I.T.I, LaTIM ERM-0102, ENST Bretagne, BP 832, 29285 Brest Cedex, France
[2]Service Echodoppler, LaTIM ERM-0102, CHRU La Cavale Blanche, 29200 Brest Cedex, France
[3]Faculty of Engineering, Multimedia University, Jalan Multimedia, 63100 Cyberjaya, Selangor, Malaysia

Abstract - **Venous thrombosis is a common pathology that creates serious problems in public health. The diagnostic of thrombosis, particularly the determination of their relative ages can be efficiently accomplished by ultrasound imaging. This study intends to classify automatically the thrombosis ages by using a predefined learning base that depends on a prior knowledge of physicians. In practice, this learning base is affected by information imperfections of the type ambiguity since physicians cannot give exact thrombosis ages. Thus, the proposed learning base is constructed in a 3-tuple: observation, label, membership value in term of fuzzy logic for each class and not a 2-tuple as in the usual supervised Bayes classification application. By considering this "fuzzy learning base", a method modeling simultaneously the concept of probabilistic uncertainty and ambiguity is proposed. In this approach, the probability for a given observation is considered on the membership value of each class and not on the class itself. At this level, the discussion focuses on two types of applications: the thrombosis ages classification and the definition of membership function by using a fuzzy learning base for classification.**

Keywords - **Venous thrombosis, echography, ultrasound imaging, age classification, fuzzy logic, probability.**

I. INTRODUCTION

The characterisation of blood clots and especially the explanation or knowledge of their maturing process, are difficult to be comprehended due to the lack of efficient analysing methods or diagnostic procedures. Echography, an indirect but simple imaging system, presents an interesting potential in terms of resolution, discrimination capabilities and 3D possibilities to analyse these pathologies.

The common encountered problems in medical imaging include segmentation, reconstruction, classification and diagnostic aid systems. Until now, there are not much works being done in the domain of supervised classification of pathologies by ultrasound imaging. This is due to certain circumstances and existing difficulties in medical as well as in image processing domain that complicate this problem. The learning base that we obtain is not perfect since the interpretation of pathologies by echographic images is not easy, as it requires some expertise experience. Furthermore, the diagnostics of pathologies given by physicians are usually qualitative. Different physicians while examining patients with the same pathologies may give different judgements. On the other hand, the extraction of features that can interpret correctly the medical criteria such as echogenecity and echostructure are perturbed by the presence of speckle noise in echographic images. It is necessary to quantify the pertinence of features and verify that they are sufficiently discriminating for the classification purposes.

In this work, a method that is able to improve the performance of a classification system is discussed. After a brief description of the medical problematic of this study in section II, the methology is presented in section III. Conclusions and future works end the discussion of this paper.

II. MEDICAL PROBLEMATIC

At the present day, some recent studies go toward the development of 3D free hand and ultrasound system [1]. This study intends to introduce a method for thrombosis age classification through the use of this kind of acquisition system. In order to achieve this purpose, we aim to implement a 3D acquisition system so that well-aligned images can be acquired and 3D analysing of data can be achieved. Until now, the developed system is a prototype that allows *in vitro* thrombosis or blood clots acquisition. However, this system will be improved in a near future so that graded *in vivo* images can be acquired during patient examinations. This prototype is based on an existing echography and completed by an electro-magnetic 3D localizer that devotes the 3D capabilities to the system. The acquisition system that we use is a product of *Advanced Technology Laboratories* (model HDI 5000).

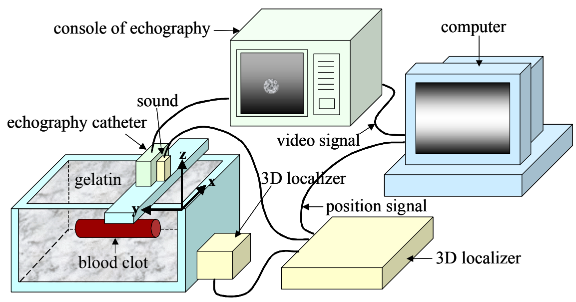

Fig. 1. A complete system for *in vitro* thrombosis or blood clot image acquisition. The 3D aligned echographic images are acquired by the movement of an echography catheter, where the acquisition position is transmitted by the sound.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle Toward Determination of Venous Thrombosis Ages by Using Fuzzy Logic and Supervised Bayes Classification	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Departement I.T.I LaTIM ERM-0102 ENST Bretagne, BP 832 29285 Brest Cedex, France	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.	
Abstract	
Subject Terms	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU
Number of Pages 4	

The study of maturing process is a fundamental issue for the medical research and particularly for the understanding of blood clots formation mechanisms. The interest from the perspective of diagnostic aid is clear. In fact, this will create the possibility of categorizing the pathologies and associating them with the corresponding therapies. Through this model, the effects of maturing process will become more comprehensible. Our interest is the classification of thrombosis age based on a series of *in vitro* thromboses images (Fig. 2).

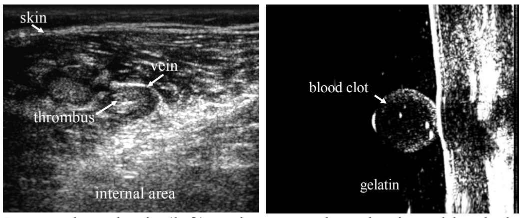

Fig. 2. *In vivo* thrombosis (left) and *in vitro* thrombosis or blood clot (right) images acquired by echography

These *in vitro* thromboses are in fact the formation of blood clot from blood sample that contributed by some volunteers. The acquisition of these blood clot images is done each day for duration of 5 days. Our purpose is to construct an example base with the collaboration of physicians since their knowledge and experience are necessary in attributing an age for each example. In a near future, the study of *in vivo* thromboses images of actual examinations will then follow (Fig. 2). For the *in vivo* case, a great number of patient examinations are required to confirm the efficiency of our studies. However, some immediate difficulties arise in controlling the example of learning base. It is affected by some imperfections due to the observation ambiguity and decision ambiguity that caused by the qualitative judgement of physians. Thus, our goal is to develop a classifier method that can take into account the ambiguity or the quality aspect of the diagnostic state attributed to the examples of learning base.

III. Methodology

In the classification problem, more generally in the pattern recognition problem, information imperfection can appear in all level: from the observed scene to the decision space (Fig. 3). These imperfections may be caused by the sensor, the feature extraction process and the decision system itself. All these imperfection sources can be modeled by the consideration of an information element (IE) as in [2]. The manner to define an IE allows us to model different kinds of information imperfections such as inaccuracy, ambiguity, uncertainty and evidential.

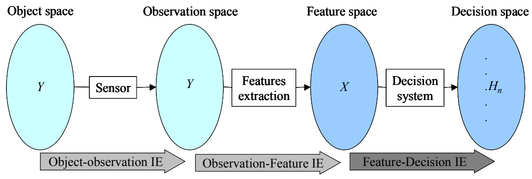

Fig. 3. A general scheme of pattern recognition system

This work focuses on the supervised classification in considering the feature space and the decision space. In actual situation, there may have ambiguity in observation, features and decision space that affect the learning base. If the observations are ambiguous, then the consideration of fuzzy propositions P_i ("A is B" where B is a given concept) is needed, whereas if the imperfection concerns the classes C_i, then we have to consider fuzzy set in the decision space (Fig. 4).

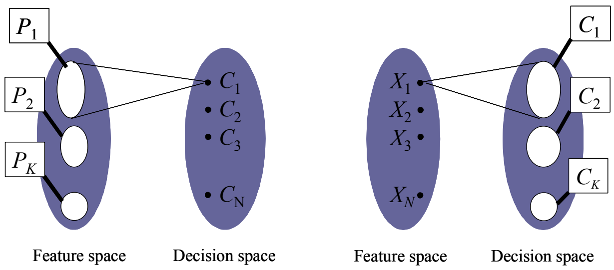

Fig. 4. Learning base possible imperfection

A. *Fuzzy concepts to model physician diagnostic ambiguity*

In this study, the decision space is affected by ambiguity since the physician cannot provide an exact thrombosis age. Due to this reason, we propose the physician to assign a "degree of truth" for each proposition of thrombosis age. By this way, a membership value is assigned to each class C_i, $i=1,\ldots,K$ for a given thrombosis case, where every class is a finite time interval. Thus, the notion of membership function (initially introduced by Zadeh [3]), relative to a class C_i for a given thrombosis case μ_{ci} is defined as follows:

$$\mu_{c_i} : \overline{X} \rightarrow [0,1]$$
$$\mu_{ci} : X=x \rightarrow \mu_{ci}(X=x) \qquad (1)$$

where $\overline{X}$ is the feature space. These class of membership function denotes the relation between the observed feature x to the concept of class C_i.

The physician reasoning is qualitative and based on their own experience. They are also against the idea to quantify the "degree of truth" on a diagnostic. As in [3], we propose to establish a scale of relation between the linguistic variables and the numerical world. The relation must be done with a particular attention since we have to consider all possible situations posed by the application. For example, we suppose that there are K classes C_i, $i=1,\ldots,K$ and the values of membership functions are interpreted as

$\mu_{c_i} = \nu$	*0*	*0.2*	*0.4*	*0.6*	*0.8*	*1.0*
degree of truth	definitely not	not very sure	not sure	sure	very sure	Definitely sure

At the same time, we cannot conceive to ask the physician to attribute for each new case some membership value of their degree of truth to each possible class. Thus, we propose the physicians to give their opinion on a thrombosis age and provide a degree of truth via linguistic variables as {"definitely not", "not very sure", "not sure", "sure", "very sure", "definitely sure"}. This idea may be processed as in Fig. 5, where the distribution of membership value to each class C_k is a gaussian that centered on the class C_k itself, with a standard deviation that is proportional to the degree of truth by the proposition of the physicians. By this way, the physicians can define easily for each thrombosis case a degree of truth.

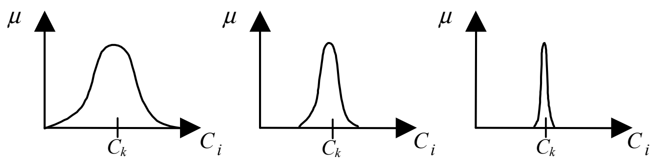

Fig. 5. Example of membership distributions due to the physician's diagnostic for a thrombosis case that being classified in the class of thrombosis age C_k, with the degrees of truth as "not very sure", "sure", "very sure" respectively.

B. *Learning base*

The learning base is constructed according to the physicians' reasoning. The decision making of a new thrombosis case is based on the expertise knowledge and the patient medical history. By using the echography, each diagnostic is compounded by an age estimation added with an intuitive "degree of truth" on this decision. Suppose that all possible classes of thrombosis ages are grouped in the set $\Omega = \{C_1, \ldots, C_K\}$, the "fuzzy learning base" is then defined as follows:

$$\mathrm{B} = \{(X, C_i, \mu_{C_i}) \mid i = 1, \ldots, K\}, \qquad (2)$$

where X is a feature vector and μ_{Ci} is the membership function to each class C_i. With this fuzzy learning base, the probability is considered on the membership function for each class C_i in order to make a decision based on both the concepts of probabilistic and fuzzy logic.

C. *Classification approach*

The Bayesian classification is based on the "closed world assumption". That means we have the entire knowledge on the component of the decision space and no other component is possible. In the context that all possible classes of thrombosis ages are grouped in the set $\Omega = \{C_1, \ldots, C_K\}$, we assume that it is possible to consider information exclusivity and exhaustivity. That means we have:

$$\sum_{i=1}^{K} P(C_i) = 1 \text{ and } \forall i \neq j,\ C_i \cap C_j = \phi . \qquad (3)$$

From now onwards, we consider the problem from the Bayesian point of view. Given a stochastic vector of features labeled $X=x$, we can express the following conditional probability:

$$P(X=x|C=C_i) = \frac{P(C=C_i|X=x)P(X=x)}{P(C=C_i)} . \qquad (4)$$

The decision-making is based on

$$P(C=C_i|X=x) = \frac{P(X=x|C=C_i)P(C=C_i)}{P(X=x)}, \qquad (5)$$

where $P(X=x)$ is the feature probability, $P(C=C_i)$ is the *a prior* probability of a class and $P(X=x|C=C_i)$ is the probability of the observation of feature X belonging to the class C_i.

This theoretical concept requires that all hypotheses are perfectly satisfied. In particular, the considered features have to be discriminant enough and all possible classes have to be perfectly and precisely known. At the same time, the learning base imprecision limits such an approach. Thus, all classes associated with the observations are perfectly known but not having the sufficient accuracy. Here, we do not associate a feature vector with a class but a feature vector with a fuzzy set to the class C_i. Works have already been done by posing some bases on how to relate fuzzy set and probability theories [4]. In our application, we establish the bridge by modifying the conditional probability defined in (4) to obtain a new definition as follows:

$$P(X = x|\mu_{ci} = \nu) = \frac{P(\mu_{ci} = \nu|X = x)P(X = x)}{P(\mu_{ci} = \nu)} . \qquad (6)$$

This leads to a new Bayes expression by taking into account the learning base imprecision:

$$P(\mu_{ci} = \nu|X = x) = \frac{P(X = x|\mu_{ci} = \nu)P(\mu_{ci} = \nu)}{P(X = x)} \qquad (7)$$

In the last relation, $P(X=x)$ is the feature probability, $P(\mu_{ci} = \nu)$ is the probability of a given membership value relative to class C_i, $P(X = x|\mu_{ci} = \nu)$ is the probability of the observation of feature X considering a degree of truth that belong to the class C_i. The expression (7) seems very similar to (5), but its interpretation is totally different. In this case, we do not evaluate the power of the link between a thrombosis case and a class, but the link between this thrombosis case with a fuzzy set associated with the membership function μ_{ci}.

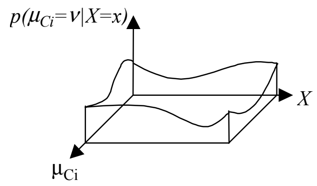

Fig. 6. For each class C_i, the probability to observe a membership value to the class C_i, given the feature X, can be represented as a surface. In this case, a two dimensions space is considered while in the classical Bayes rule, only a mono-dimension space is considered.

The understanding cannot be perfect without being sensible to the following considerations and in particular of the consistence of μ_{Ci}.

- The cases with different features but have the same membership values: Two patients with different thrombosis maturing state might have been assigned the same degree of truth to certain class by the physician, that is,

$$x_i \neq x_j \text{ but } \mu_{ck}(x_i)=\mu_{ck}(x_j). \qquad (8)$$

- The cases with the same feature but have different membership values: Two patients with the same thrombosis maturing state might have been assigned different degree of truth to certain class by the physician, that is,

$$x_i = x_j \text{ but } \mu_{ck}(x_i)\neq\mu_{ck}(x_j). \qquad (9)$$

These two remarks show how difficult it is to establish a perfect relation between the classes and the observation, by using only fuzzy logic. That is why the probabilistic point of view is very useful. Now, we can develop the following criteria in some simple particular cases in order to establish a possible approach for the decision-making process:

1) If the value of membership function $\mu_{ci}=v$ is high and the probability $P(\mu_{ci}=v\,|\,X=x)$ is high, then the probability of the observation in class C_i is high.
2) If the value of membership function $\mu_{ci}=v$ is low and the probability $P(\mu_{ci}=v\,|\,X=x)$ is low, then the probability of the observation in class C_i is high.
3) If the value of membership function $\mu_{ci}=v$ is low and the probability $P(\mu_{ci}=v\,|\,X=x)$ is high, then the probability of the observation in class C_i is low.
4) If the value of membership function $\mu_{ci}=v$ is high and the probability $P(\mu_{ci}=v\,|\,X=x)$ is low, then the probability of the observation in class C_i is low.

It is important to understand that on real data, the problem of decision-making is harder and need to be studied more precisely when ambiguous cases are encountered. For example, it is difficult to make decision when the membership value $\mu_{ci}=v \approx 0.5$ and the probability $P(\mu_{ci}=v\,|\,X=x)\approx 0.5$. Also, the problem of ambiguity may arise in decision-making when the membership values and the probabilities for two different classes C_i and C_j are very close, that is, $\mu_{ci}=v_i\approx v_j=\mu_{cj}$ and $P(\mu_{ci}=v\,|\,X=x)\approx P(\mu_{cj}=v\,|\,X=x)$.

IV. CONCLUSION

This preliminary study proposes to develop an acquisition system and some methods to classify pathologies from ultrasound image sequences through the consideration of 2D and 3D features. More particularly, we deal with a particular type of imperfection of the kind of ambiguity, which can alter the quality of learning bases when we focus on supervised classification methods. The considered application i.e. the thrombosis age evaluation, enter exactly in this kind of problematic. These considerations, which established with a very closed participation of physicians, have permitted to pose an adapted protocol of example bases establishment. From this point of view, the idea to exploit the learning base as well as the possibility in modeling its imperfection has been discussed. We have briefly proposed the integration of this kind of imperfection with a classical Bayes supervised classification approach. We can conclude that the idea to combine the management of both the uncertainty and the information ambiguity is a very interesting challenge for an academic as well as an applicative point of view in the near future. The application of thrombosis analysis shows that it is necessary to consider the probabilistic and fuzzy logic approaches not just as concurrent methods but like two complementary way to process the information. Further works will clarify decision-making and compare the classical classification methods with this proposed method in the case where the learning base is affected by ambiguity.

REFERENCES

[1] Barry C., Allot C., John N., Melor P. Arundel P., « Tree Dimensional Free Hand Ultrasound: Image Reconstruction and Volume analysis » Ultrasound in Med. & Biol., vol. 23, n°9, pp.1209-1224, 1997.
[2] Basel Solaiman, « Information fusion concepts from information elements definition to the application of fusion approaches », Proceeding of the SPIE Aerosense conference, vol. 4385, pp.205-212, Orlando, 2001.
[3] L.A. Zadeh., « Fuzzy sets. Inform ». Inform. Contr., vol. 8, pp. 338-353, 1965.
[4] D. Dubois, H. Prade, « Fuzzy sets and probability: misunderstandings, bridges and gaps », Proceedings of the second IEEE International Fuzzy Systems conference, pp. 1059-1068, vol.2, 1993.