Bayesian Classifier for Medical Data from Doppler Unit

J. Málek

Nowadays, hand-held ultrasonic Doppler units (probes) are often used for noninvasive screening of atherosclerosis in the arteries of the lower limbs. The mean velocity of blood flow in time and blood pressures are measured on several positions on each lower limb. By listening to the acoustic signal generated by the device or by reading the signal displayed on screen, a specialist can detect peripheral arterial disease (PAD).

This project aims to design software that will be able to analyze data from such a device and classify it into several diagnostic classes. At the Department of Functional Diagnostics at the Regional Hospital in Liberec a database of several hundreds signals was collected. In cooperation with the specialist, the signals were manually classified into four classes. For each class, selected signal features were extracted and then used for training a Bayesian classifier. Another set of signals was used for evaluating and optimizing the parameters of the classifier. Slightly above 84 % of successfully recognized diagnostic states, was recently achieved on the test data.

Keywords: Medical data recognition, hand-held ultrasonic Doppler unit, peripheral arterial disease.

1 Methods

The signals for classification were measured by a hand-held Multi Dopplex II ultrasonic unit and sent to a PC for storage via an RS232 interface. The device measures the mean velocity of blood in an artery within a short time period, a Doppler velocity waveform.

Multi Dopplex II is a bi-directional device. The waveforms can be displayed as forward and backward flow (see Figs. 1–3) or as the difference of these two signals in combined waveforms.

In a standard situation, five positions on each leg are examined: artery femoralis, a. poplitea, a. tibialis posterior, a. tibialis anterior and a. dorsalis pedis.

For automatic classification, four classes were chosen, into which the signals will be classified. These classes reflect various degrees of artery occlusion, and also describe some defects that can be considered by a specialist during this examination.

1.1 Classes

Normal course – Signals acquired by an examination of arteries without peripheral arterial disease (PAD) – Fig. 1.

Stenotic course – Signals measured in arteries with a stenotic diameter – Fig. 2.

Occlusion – Signals measured in arteries with total arterial obstruction – Fig. 3.

Incorrect course – Various errors may occur while measuring e.g.: the amplification factor was set too high – the course is cut; the signal is under strong influence from nearby veins, and so on.

1.2 Features

During the design process, 20 features were considered as potentially useful for the classifier. These features describe the quality of the measured signals in the time domain, in the frequency domain, or have a special medical meaning. The Sequential Forward Search (SFS) algorithm was used to deter-
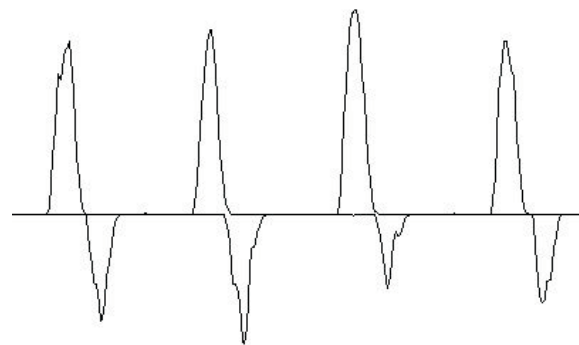

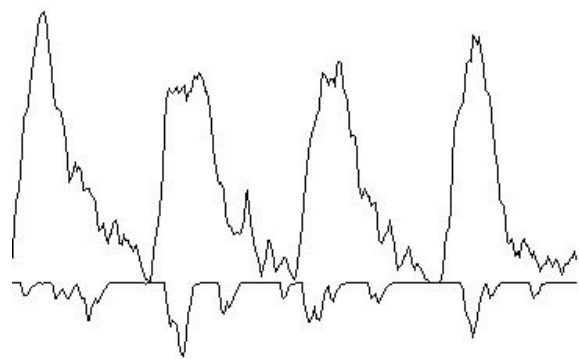Fig. 1: Signals acquired in an examination of arteries without peripheral arterial disease (PAD)

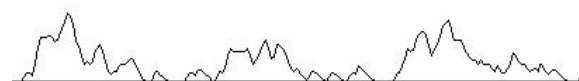Fig. 2: Signals measured in arteries with a stenotic diameter

Fig. 3: Signals measured in arteries with nearly total arterial obstruction

mine the most significant features. It was determined that these were the features usually considered by medical specialists during the visual examination process. SFS found the following four features:

Signal energy – Log sum of squared signal samples over the whole signal.

Velocity/time index – Difference of maximum and minimum blood velocity within one pulse region divided by the pulse region duration.

Deceleration – Maximum velocity within one pulse region divided by its fall time.

Brachial Pressure Index (BPI) – Ratio of the distal pressure in an examined position on the lower limb and the patient's system blood pressure (measured on a. brachialis).

Before computing the features, the signal must be preprocessed. This was done by filtering the signal by a low pass filter to suppress the high frequency noise.

It should be noted that all the features are calculated automatically in real time, without human intervention.

1.3 Bayesian classifier

The Bayesian classifier represents each of the four classes by one or more Gaussians (modes) within the probability density function (pdf) in a 4-feature space. Their parameters (means and variances) are estimated from the training data. Since the classes have different occurrence rates, the class prior probabilities are also taken into account. The number of Gaussians within one class depends on the distribution of the features in 4-dimensional space and is determined by a vector quantization process.

The database uses data from 1,100 examinations, measured in 10 standard positions (5 on each leg), i.e. 11,000 samples are available. In the experiments, data from 880 randomly chosen examinations was used for training the classifier; the remaining data (from 220 examinations) was left for testing. This was repeated 5 times in order to obtain more reliable results. The classifier assigned to each signal the class with the highest posterior probability. The result was compared with the expert's decision. After this the recognition scores were calculated for each of the measuring positions and then averaged.

2 Results

The optimal design of the classifier was investigated in many experiments. The most relevant results are shown in Table 1. It is evident that the best performance was achieved by the classifier with multi-modal pdfs and prior probabilities. The best score was 84 %, which means that in 84 % of the cases the classifier and the expert agreed.

Table 1: Results for different types of classifier

Method	Score [%]
1-modal pdf without prior probability	78.66
1-modal pdf with prior probability	80.22
Multi-modal pdf without prior probability	81.95
Multi-modal pdf with prior probability	84.01

3 Discussion

The recent version of the program is based on probabilistic methods. In an attempt to improve the models, I intend to adopt the existing training strategies for so called Gaussian Mixture Models (GMMs) that have already been developed in the Speech Lab at TU of Liberec for speech signal classification.

In the next step I intend to focus on implementing better frequency features calculated for the combined flow waveform rather than for separated forward and backward flow. It will be necessary to make the peak detection and feature calculation more precise. Unfortunately, the presence of vein signals in the data and noise in the low amplitude signals complicates this task. It also seems useful to extend the number of classes for more precise classification of the signals. The class Stenosis may be divided into two subclasses: mild and severe Stenosis.

4 Conclusions

The method investigated here shows possible new ways to perform screening for PAD in general practice surgeries. Computer analysis and classification could help to achieve an early diagnosis of grave vascular diseases, so that treatment can be given before the development of critical limb ischemia.

5 Acknowledgments

References

[1] Transatlantic Inter-Society Consensus (TASC), *International Angiology 2000*, Management of Peripheral Arterial Disease (PAD); Suppl. 1.

[2] Wolfe, J. H. N. et al.: ABC of Vascular Diseases. *BMJ* 1992, p. 12–14.

[3] Topol, E. J. et al.: *Atlas of Atherothrombosis*, London: Science Press; 2004.

[4] Goldman, L. et al.: *Textbook of Medicine*, 22nd ed., Philadelphia: Saunders, 2004.

[5] Rozman, J. Lékař a technika, Praha: NTS ČLS *J.E.P.* 1995, p. 9–11.

Jiří Málek
e-mail: Jiri.Malek@tul.cz

Faculty of Mechatronics and Interdisciplinary
Engineering Studies
Technical University of Liberec
Hálkova 6
461 17 Liberec, Czech Republic